Assessing research performance in UK universities using the case of the economics and econometrics unit of assessment in the 1992–2014 research evaluation exercises

Johnston, James; Reeves, Alan

Published in:
Research Evaluation

DOI:
10.1093/reseval/rvw021

Published: 09/01/2017

Document Version
Peer reviewed version

Assessing research performance in UK universities using the case of the Economics and Econometrics unit of assessment in the 1992 to 2014 research evaluation exercises

James Johnston and Alan Reeves [1]

Abstract

Research evaluation exercises (REEs) affect the allocation of research funds both within and between universities. How the results of REEs might be used by university managers in decisions on which areas of research to support is the central focus of this paper. The decision on whether to support research in an area is explained by reference to an institutional threshold level, defined here as the minimum acceptable research score. Data from submissions to the Economics and Econometrics (E&E) unit of assessment (UOA) in the various UK research evaluation exercises appear to support the predictions of the model. Two types of gap are defined – internal and external. Negative internal or external gaps are found to be closely related to the decision to withdraw from the E&E UOA at the next REE. The information is being used by universities in ways that appear to have had far-reaching consequences for research and researchers.

Keywords: Research evaluation exercises, MARS, Economics and Econometrics unit of assessment, internal and external gaps, withdrawal of research support.

[1] James Johnston and Alan Reeves

Introduction

Research evaluation exercise (REE) scores have two distinct groups of users: first, the state which funds research and second, universities, which produce research. Research funders need a basis on which to allocate their funds. Prior to formal research evaluation state support for research was largely distributed on an informal basis (Martin and Whitley, 2010). Commentators such as Jarrat (1985) criticised what they saw as too cosy an arrangement and argued that universities should operate more like businesses. The next few decades saw all aspects of the university – including its approach to the management of research - transformed. A system of formal Research evaluation exercises (REEs) gradually developed. Both lauded and denounced REEs 1986 to 2014 have become an important feature of life in UK universities (Bence and Oppenheim, 2005; Sharp, 2005; Harley and Lee, 2007). Some fourteen countries have followed the UK's lead by introducing performance-based research funding systems (Hicks, 2012). The proportion of research funds distributed on the basis of REE ratings has risen from 17% in 1986 to over 90%. (Lee, Pham and Gu, 2012; Leisyte and Westerheijden, 2014). High research ratings may mean more funds and low ratings may mean no funding at all, putting at jeopardy whole departments and academic jobs (Field, 2015). A survey by the University and College Union (UCU), a union for academics in the UK, found that many of its members feared that a negative internal judgement of their research performance could lead to them being transferred onto teaching-only contracts or denied promotion (Owens, 2013). In addition to financial benefits a favourable REE evaluation may boost individual and institutional reputational capital. Indeed, Hicks (2012) concluded that it is the desire for academic prestige that is most important in stimulating improvements in the quality of research.

REEs attempt to capture the complexity and diversity of research in a number that can then be used to rank the research performance of individuals and the teams and institutions of which they are a part. The independent experts that make up REE assessment panels inform the uninformed about the quality of research output. This process comes at a cost and the high cost of REEs in the UK has led some observers to suggest the use of systems such as those operating in parts of Europe (Guena and Martin, 2003). One interpretation of the fact that scarce resources continue to be willingly expended on REEs is that those that fund REEs take the view that these exercises yield sufficient offsetting benefits in the form of higher quality information on research performance and that this in turn may lead to improved allocative efficiency. Supporters of REEs argue that the greater competition between universities for funding as a result of REEs has led to an increase in the quality-adjusted amount of research produced by the higher education system as a whole (Curran, 2000; Lucas, 2006; Thomas, 2007). Elite institutions themselves maintain that the concentration of scarce funding on research that is genuinely 'world class' is what is required to enable the UK's elite institutions to compete with rivals in other parts of the world (Russell Group, 2009).

There are three broad strands of criticism of REEs. First, for some, REEs are antithetical to the whole ethos of higher education (Willmott, 2011). Questions have been raised about the extent to which it is possible to secure research commensurability even within a narrowly defined field through a process that reduces diverse research to a single metric (Agyemang and Broadbent, 2015; Bence and Oppenheim, 2005). Others, such as Willmott (2011) and Mingers and Willmott (2013), maintain that REEs have led to the Taylorization of research in

universities and that this has been detrimental to the pursuit of long-term and risky research. The second broad type of criticism reflects the view that they may have undesirable unintended consequences. One of the dangers of a system of research funding in which current funding is directly proportional to assessments of earlier research performance is that it risks entrenching elitism (a 'Matthew effect') and so works against open competition among researchers (Merton, 1968). How does this work in practice? Institutions with the best research scores get the most funds and providing elite institutions with a financial advantage over middle-ranking institutions. Elite institutions can then poach outstanding staff from middle-ranking institutions and they in turn will do the same to the lower-ranking institutions. This behaviour strengthens the elite and weakens lower-ranking rivals and is a well-known feature of the academic system, particularly in the run up to a REE when the incidence of job-switching among academic staff intensifies. High performers are better able to attract, motivate and retain the best academic talent. Financial strength also enables institutions to invest in the skills of research students and assistants, who will in turn become the next generation of top researchers. The greater the focus on research excellence in a funding system the more intense this dynamic will be. Lee (2007) and Lee, Pham and Gu (2013) have shown that the approach to research evaluation in the UK has acted to fundamentally reshape subjects such as economics, with non-mainstream approaches being driven even further to the margins of the subject. Rafols et. al. (2012) found that the use of journal rankings in REEs may work against 'academically and socially useful' interdisciplinary research. Recent literature has raised the important issue of performativity and the reactive effects of the REEs (Sato and Endo, 2014). Finally, a third strand of criticism is based on the belief that while evaluation has the potential to be helpful, the way that it is carried out is problematic for a variety of reasons. (Bence and Oppenheim, 2005; Sato and Endo, 2014; Leisyte and Westerheijden, 2014).

This paper is primarily concerned with how research producers such as universities and other higher education institutions make use of the information provided by REEs. It examines possible institutional research goals and seeks to explain university reactions to REE scores with reference to the concept of the minimum acceptable research score (MARS). A model of research performance involving two new concepts – internal and external gaps - is developed and tested in this paper. The model allows a university to compare its performance in a particular research area (or unit of assessment (UOA)) with other research areas in the same university (the internal element); and with that same research area in other universities (the external element). Data on the Economics and Econometrics (E&E) UOA at the various REEs are used to test the predictions of the model. The model turns out to be a good predictor of whether a university continues to the next REE or withdraws and devotes resources to other areas of research. It also emerges that the external comparison is more important than the internal comparison in explaining the research support decisions of universities.

University goals and the internal allocation of research funding

University mission statements often proclaim that they will only do what they are good at doing. To help ensure that this happens, institutions carry out internal portfolio reviews to assess the health or otherwise of its various subject areas. Factors such as the strength of student demand for programmes, the quality of the staff involved and assessments of research strength will all be important in whether universities continue to support research or

to offer programmes in an area. It seems likely that these factors may be traded-off against one another, e.g. the strength of student demand for programmes may be traded-off against its research performance. While a university grouping that struggles to attract well-qualified students in sufficient numbers may be retained as long as the group is strong in research, its programme might be withdrawn and its resources redeployed or released if research performance is weak. Academics retained on the basis of good anticipated research scores may find their positions under threat if REE scores fall below what is deemed to be the minimum acceptable level. Fewer resources or a greater desire for productive efficiency in the use of resources on the part of university managers may manifest themselves in a greater willingness to move resources from poor to strong performers. Poor performers will be 'weeded out' and resources will be concentrated on a smaller number of stronger areas. The longer this reorganisation takes the greater the cost incurred by the institution.

The precise research goals pursued by universities are often opaque or, if not, remain known to a small and closed group, despite many universities publishing their research strategies. Still, university managers always have the option of shifting resources from one academic area to another. Each area can therefore be thought of as engaging in internal competition for funds. It is instructive to consider the possible research aims of universities and how these might determine institutional reactions to REE scores. One extreme possibility is that a university seeks to maximise its research score irrespective of the number of UOA submissions. The pursuit of such a narrow goal would imply that any UOA that scored below an institution's highest score would no longer be supported, with research becoming concentrated in a smaller number of areas. Such a university would gradually move to a smaller number of high performing units. In this scenario if a university achieved the same score in all UOAs there would there be no incentive to engage in post-REE reorganisation of research support. It is worth emphasising that this prediction regarding the reaction of universities to REE scores would apply in universities where research scores were uniform, regardless of whether they were uniformly good, moderate or poor, less post-REE reorganisation would be expected. In any post-REE reorganisation of research, university managers would have to weigh adjustment costs (monetary and non-monetary) against benefits in the form of higher levels of research output. Monetary costs might include human resource costs (e.g. redundancy or retraining) and the loss of teaching-related revenues if courses are withdrawn. Non-monetary costs might include the harm to the welfare of staff adversely affected and the potential damage to the institution's reputation. Higher adjustment costs would lead to less and slower reorganisation.

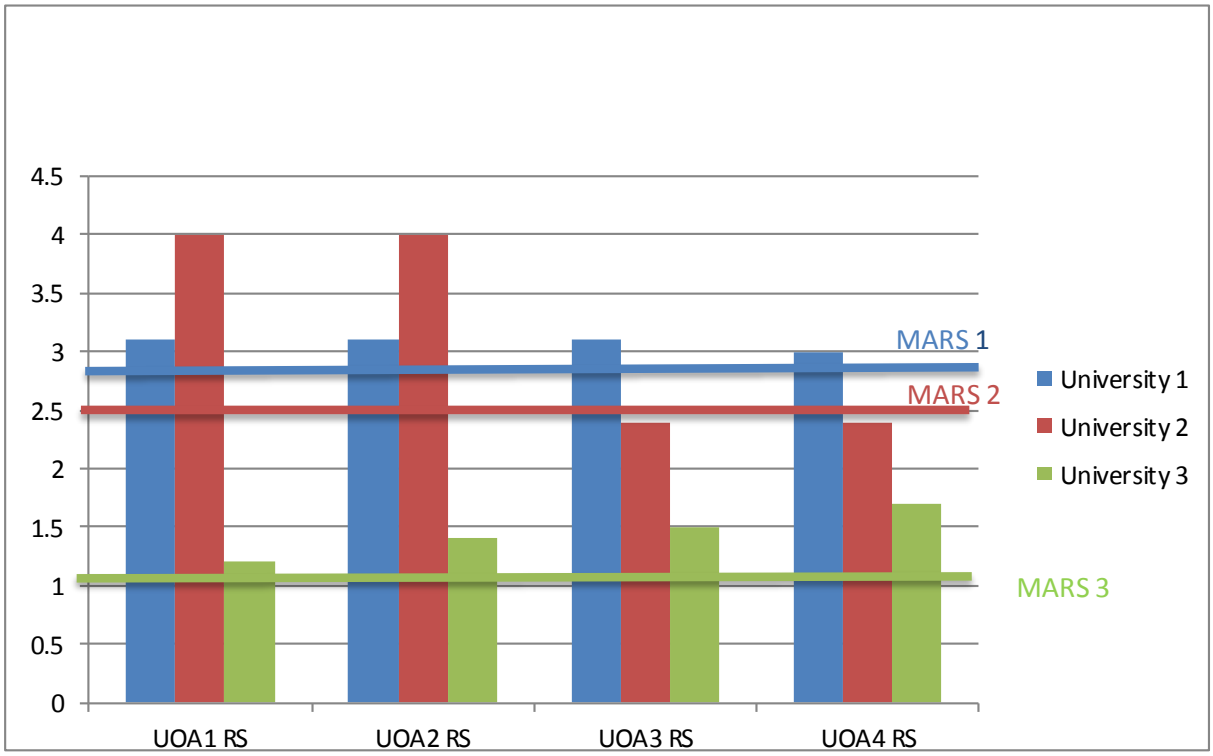
A less knife-edge approach might involve universities withdrawing support to UOAs that score below the institution's average score for all UOAs. It is important to note that university managers as well as making internal comparisons are likely to also make external comparisons with their rivals. So it is not just how a UOA has performed relative to its internal peers but how it has done when compared to other institutions that will be important to university managers. That said, once again institutions that are uniform performers would not be expected to reorganise as much as those with high levels of variation in performance. The more incremental approach to change implied by this objective would allow university managers to navigate the development of their organisations without the upheaval that might accompany a focus on excellence at all costs. The adjustment costs associated with this approach to research management would be less pronounced but would need to be balanced against the cost of lost research output that would result from not moving resource

to more productive areas. A university that adopts this gradual approach would lose out if rivals were to adopt a more efficiency-based approach.

Research scores, minimum acceptable scores (MARS) and internal funding

When deciding on whether to support an area of research a university must compare the area's actual REE score in a UOA against some sort of threshold level or MARS. For example as part its 'Raising the Bar' initiative research-active staff at Newcastle University were given 'minimum expectations for research performance' (Grove, 2015). This can be thought of as analogous to the required rate of return on capital for a for-profit business or a hurdle that research groups have to jump to ensure continued support for their activities. At any given time the MARS may vary between institutions and within institutions it may vary over time. It may take the form of an institution-wide or unit-specific minimum. Institutions may make its level known to research staff in advance of a REE in which case it may come to act as a target for researchers. With respect to its value factors such as an institution's performance in earlier REEs, both internally and externally, the resource expended on the recruitment of new staff and internal politics are, inter alia, likely to be important in determining the MARS. Research activity in traditional disciplines may be regarded by some elite university mission groups as part of the 'price' of club membership. If this is the case then institutions may continue to support research in an area even though it has performed relatively poorly. Universities that have made non-transferable subject-specific investments in research areas may be reluctant to close down research in that area: the larger and more specific the investment the less likely a university is to withdraw for any given level of performance.

Figure 1: REE scores (RS), minimum acceptable research scores (MARS) and research support decisions



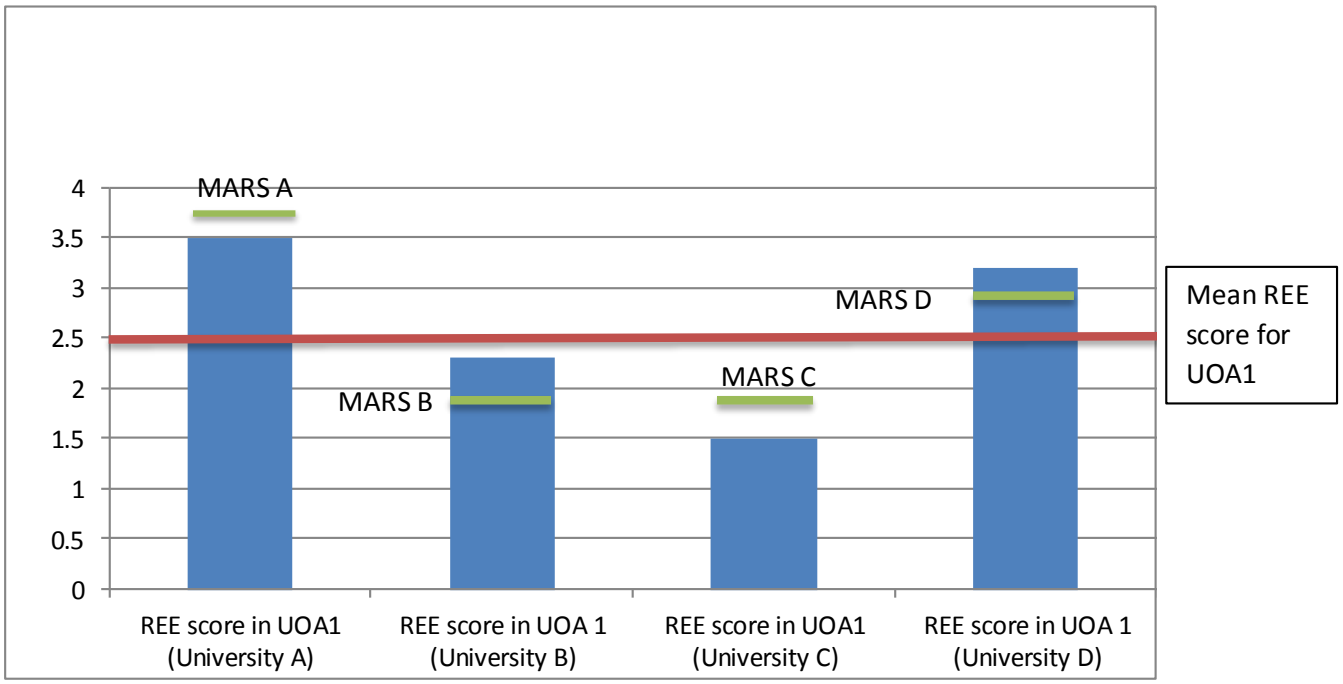
With regard to internal politics, strong researchers may lobby university management for a higher institutional MARS. In contrast, those who are either less good or confident about their prospects will push for a lower threshold so as to avoid the removal of support. Some groups may even push for a group-specific MARS set at a lower level than that for other groups in the same institution. However, as research funds can always be switched between different areas an institution-specific MARS is likely to hold sway over any attempt to entrench inefficiency through the adoption of unit-specific MARS.

Figure 1 illustrates how the MARS pursued by an organisation might be expected to condition its decisions on internal funding allocation. University 1 has the highest institutional MARS of the three institutions and finds that its submissions in all units exceed this level. If meeting the institutional MARS is sufficient to ensure continued funding for a research area, University 1 would be expected to make few if any changes to its research portfolio. In contrast University 2 has an intermediate level MARS and finds that two of its submissions have a research score below its threshold. If achieving the MARS is a necessary condition for continued support, then two of University 2's submissions will no longer be supported. University 3 has the lowest MARS but all three submissions exceed the target and so we would expect no change in this institution's research portfolio. It is worth noting that the university that is predicted to change its portfolio most is University 2 and that this happens even though it has the highest average score of all three universities. Two of the units supported by University 2 would have their support withdrawn and this would be diverted into increased support for UOAs 1 and 2.

REE scores, the institutional MARS, internal and external gaps

A university may compare its performance in a particular UOA with other research groups in the university (the internal element). Equally it may compare its UOA score with the average of all submissions to a UOA (the external element). Which of the two comparisons is most important is not clear on a priori grounds.

Figure 2: REE scores, the institutional MARS and internal and external gaps



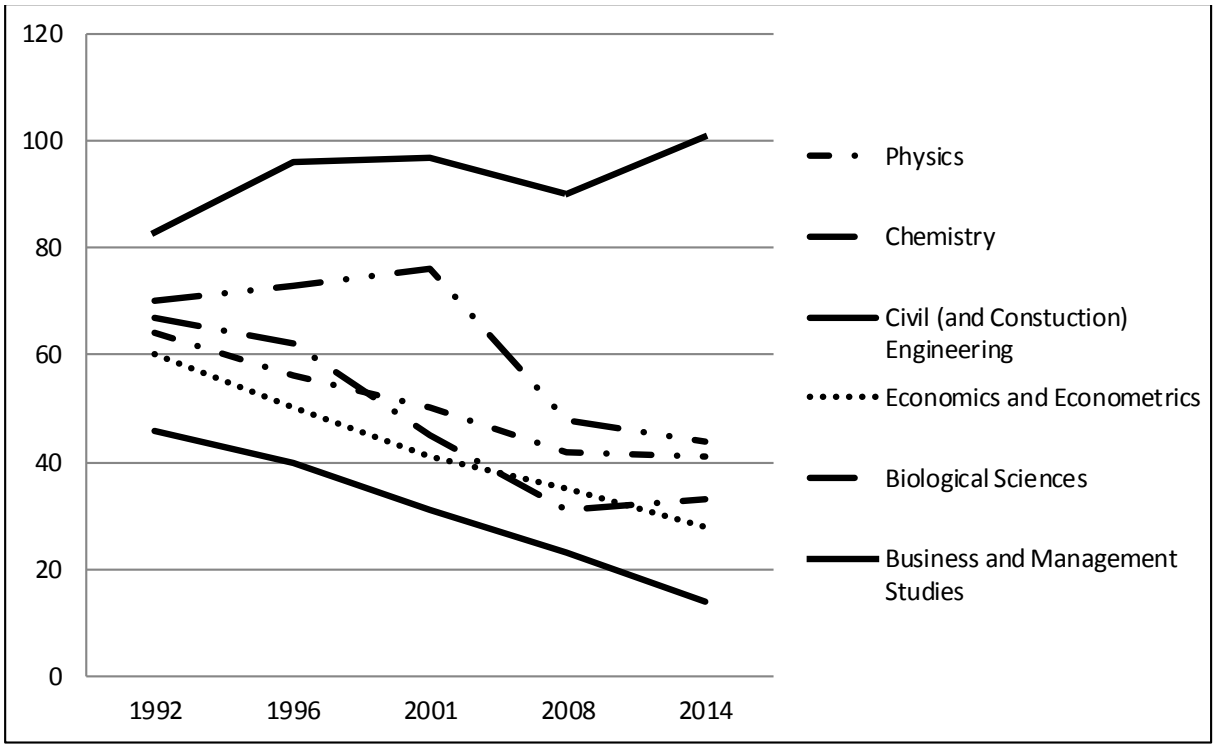
The model throws up four possibilities that are illustrated in Figure 2. University A, despite having the highest score in UOA1 has a positive external gap but a negative internal gap, University B has a positive internal but negative external cap, University C a double negative external gap and University D a double positive gap. If these gaps are important in the decision to withdraw from a UOA for the next REE we would expect universities such as University C with a double negative to be most likely to withdraw. Those most likely to continue research in an area would be those that enjoy a double positive. Where there is a positive and negative gap the picture is less clear. University A is clearly the best performer but its performance may be poorer than that of others areas within the same institution and so it may be deemed to be underperforming and at risk. Although University B has failed to achieve the mean score for UOA1 it may have performed well in comparison to other areas in the same institution and so be viewed as high performer. It would certainly be at less risk than University C with which it shares a MARS. In this institution research in UOA1 is poor compared to the unit as a whole but also in comparison with other areas in the same institution.

Data and methods

The extent to which the decision to continue to support or to withdraw from a research area depend on the sizes and signs of internal and external gaps can only be answered by reference to the data. Figure 3 shows the numbers of institutions submitting to selected UOAs in the five REEs. This paper focuses on the E&E UOA and decline in the number of universities submitting is clearly apparent in Figure 3. With about seventy UOAs (before REF 2014) and nearly one hundred and thirty universities, analysing all UOAs in depth is impractical, hence the focus on one UOA. However even though E&E has suffered a major decline in submissions it does not differ significantly from some other in this regard, notably four STEM subjects, Physics, Chemistry, Biological Sciences and Civil (and Construction from 2014) Engineering which have all experienced large falls. In contrast, Business and Management (B&M) shows a completely different pattern with a rise over time. The possibility that much of the work done by economists may be incorporated into other UOAs such as B&M and that Figure 3 gives an overly gloomy portrait of what has happened to E&E is acknowledged.

By 2014 no new universities submitted to the E&E UOA. Similar stories can be told of Chemistry, where all 19 new universities that entered in 1992 had dropped out by 2014 and in Physics where only three new universities were entered in 2014, 16 having dropped out since 1992. In Biological Sciences in 1992, 29 out of 70 were new universities and by 2014 it was just 7 out of 44. In Civil Engineering in 1992 there were 15 new among 43 entrants but by 2014 there was only one among just 14 entrants. In all of these subjects the traditional research intensive institutions were more likely to survive the course while others have fallen by the wayside or put their efforts into other subject areas. In E&E, new universities have been home to a more heterodox approach to the study of economics (Lee et. al., 2013) and are the primary conduit through which under-represented groups access the higher education system (Johnston et.al., 2014), so this steep decline should be of concern to the UK economics community.

Figure 3: The number of universities submitting to selected UOAs 1992-2014



Because there is no information on the size of the MARS a proxy is required. An average - the $FTE\ weighted\ grade\ point\ average$ - is selected for this purpose. The GPA, or grade point average, is the standard measure of research quality used in the REEs in the UK. But it takes no account of the number of researchers submitted so it is weighted by the number of full-time equivalent research staff ($FTEs$).

For each university submitting to the E&E UOA, for all REEs, a GPA was obtained. $GPAs$ are published in RAE results publications from 1992 to 2008; and for the REF in 2014. The GPA is calculated by multiplying the percentage of research in each category by its rating (1-4), summing and dividing by 100. An $institutional\ FTE\ weighted\ GPA$ is then calculated by multiplying the GPA for each UOA to which the institution submitted by the number of $FTEs$ submitted to that UOA, summing and dividing by the total number of $FTEs$ submitted by the institution:

$$Institutional\ FTE\ weighted\ GPA = \sum_{i=1}^{m} GPAij * FTEij / \sum_{i=1}^{m} FTEij$$

where m is the number of UOAs, $GPAij$ is the REE score of the ith UOA ($i = 1 \ldots . m$) for the jth university and $FTEij$ is the number of full-time equivalent research staff entered in the submission. A similar method is used to compute the GPA for the E&E UOA, recognising the need to weight the scores by the number of $FTEs$ submitted in the E&E UOA by each university. This figure varies from more than 80 to under 10 with means per university of 17.5 in 1992 rising to 27 in 2014. With a few notable exceptions universities have submitted more staff to later REEs. The formula is:

$$FTE\ weighted\ mean\ ECON\ GPA = \sum_{j=1}^{n} ECON\ GPAj * ECON\ FTEj / \sum_{j=1}^{n} ECON\ FTEj$$

where n is the number of universities, $ECON\ GPAj$ is the GPA of E&E UOA for the jth university and $ECON\ FTEj$ is the number of $FTEs$ submitted to the E&E UOA in the jth university. As before, the sum of the products is divided by the number of $FTEs$ to give the $FTE\ weighted\ mean\ ECON\ GPA$.

To compute the internal gaps the $institutional\ FTE\ weighted\ GPA$ is subtracted from the $ECON\ GPA$.

$$Internal\ Gap = \ ECON\ GPAj - Institutional\ FTE\ weighted\ GPAj$$

The internal gap will be positive when the E&E score is higher than the overall university score, negative when it is lower and zero when is the same. For example, a gap of +0.32 indicates that the E&E score was 0.32 higher than the average score for all UOAs in that university. Conversely, an internal gap of -0.32 would indicate that E&E was 0.32 less than the overall average in that university.

The external gaps are found by subtracting the $mean\ FTE\ weighted\ ECON\ GPA$ from the $ECON\ GPA$:

$$External\ Gap = \ ECON\ GPAj - FTE\ weighted\ mean\ ECON\ GPA$$

A positive external gap signifies that a university is doing better than the average of all universities and a negative gap that it is doing worse.

Changes made over the years to the rating system presented an issue. The move from a 1, 2, 3, 4, 5 scale in 1992 to a 1, 2, 3a, 3b, 4, 5, 5* scale in 1996 and 2001 was handled by using a five point scale for the 1992 data and a seven point scale for the 1996 and 2001 data (as in Sharp, 2004). A weighted average of the quality profile was used for the 2008 RAE and the 2014 REF. In REF2014 the proportion of outputs at 3* and 4* is used as a measure of quality but this does not allow comparison with previous REEs so is not used in this paper. In all cases the score in E&E UOA could be compared directly with that in the university as a whole. The average scores computed here are not the same as the MARS described earlier. The MARS, if it is actually specified, is only known to the individual universities and may not be in the public domain, unlike the research scores. As such, the MARS is an *ex ante* concept whereas the actual scores are *ex post*. If the MARS and the average are the same it implies that any score below the average is unacceptable and vice versa. The MARS may be greater or less than the average depending on the circumstances in individual universities. In a successful research intensive university a score of say, 5 in the E&E UOA (on the 1996 and 2001 scale of 1-7), may be below the MARS (and the overall university average) and therefore be unacceptable. By way of contrast, in a new university with no record of high quality economics research a score of 5 would be outstanding and likely to be well above the overall university average and the MARS.

It is hypothesised that the signs and sizes of the internal and external gaps determine in part whether a university enters or withdraws from the E&E UOA at the next REE. A double negative gap is more likely to lead to withdrawal from the next E&E UOA while a double

positive gap is more likely to lead to entry in the next E&E UOA. There are also two other possibilities: (i) a negative internal gap and a positive external gap; and (ii) a positive internal gap and a negative internal gap. The outcome of a university having a single negative gap will depend on whether it is internal or external and whether the university is more sensitive in terms of decision making to internal or external gaps.

It may also be that some universities do not act on the basis of a single REE but take into account the scores over two or three REEs. For various reasons, such as wanting to have a very large research portfolio, some universities with negative gaps may continue to support underperforming research groups by entering them in the next REE and some with positive gaps will withdraw in order to focus research efforts in different areas as pointed out earlier.

The results section will show the gaps on a four quadrant graph which incorporates both positive and negative values of the two gaps. Each of the four quadrants represents different combinations of positive and negative internal and external gaps. In addition, a probit analysis is conducted using a model of the gaps to explore the decision on whether to enter or leave the next REE based on the signs of the gaps.

Results

Since 1992 there have been five REEs with a varying number of UOAs. In the first four REEs in 1992, 1996, 2001 and 2008 there were, respectively 72, 69, 67 and 67 UOAs. In 2014 this was reduced to 36 by combining many UOAs. The full picture of entrants to the E&E UOA is shown in Table 1 which lists the universities that submitted to the E&E UOA over the five REEs. It is clear from this that the number has fallen steadily over the course of the five REEs with just 25 universities having submitted in every REE. In all, 37 universities withdrew from the E&E UOA at some point or other. Three universities - Brunel, Kingston and Durham – withdrew but re-entered at a later date, with the latter two withdrawing again. Royal Holloway and Sheffield were late entrants with their first entry in 2001. An interesting feature is the disappearance of the new (post-1992) universities from the picture in 2014. New universities are not research intensive like many old universities and over the course of the five REEs thirteen new universities dropped out so that in the 2014 REF not a single new university submitted to the E&E UOA. It should also be pointed out that in the UK there are now more new universities than old following several waves of new university creation beginning in 1992, so the lack of a single new university in the 2014 E&E UOA is suggestive of economics research becoming more and more the metier of the elite institutions of the UK. That being said, the old universities did not have a free run. Over the same period 21 old universities withdrew from the E&E UOA so the decline in the number submitting is not due solely to new universities.

Table 1: Universities submitting to the E&E UOA 1992-2014*

	1992	1996	2001	2008	2104		1992	1996	2001	2008	2014
UCL	√	√	√	√	√	**London Met.**	√	√	√	√	x
LSE	√	√	√	√	√	**Kingston**	√	x	x	√	x
Oxford	√	√	√	√	√	**Manchester Met.**	√	√	√	√	x
Cambridge	√	√	√	√	√	**Stirling**	√	√	√	√	x
Warwick	√	√	√	√	√	**Bath**	√	√		x	x
Bristol	√	√	√	√	√	**De Montfort**	√	√	x	x	x
Essex	√	√	√	√	√	**East London**	√	√	√	x	x
Royal Holloway	x	x	√	√	√	**Hull**	√	√	x	x	x
Nottingham	√	√	√	√	√	**Keele**	√	√	√	x	x
East Anglia	√	√	√	√	√	**Liverpool**	√	√	√	x	x
Edinburgh	√	√	√	√	√	**Newcastle**	√	√	√	x	x
Surrey	√	√	√	√	√	**Northumbria**	√	√	√	x	x
Queen Mary	√	√	√	√	√	**Nottingham Trent**	√	√	x	x	x
York	√	√	√	√	√	**SOAS**	√	√	x	x	x
St Andrews	√	√	√	√	√	**Portsmouth**	√	√	x	x	x
Manchester	√	√	√	√	√	**Reading**	√	√	x	x	x
Glasgow	√	√	√	√	√	**Salford**	√	√	x	x	x
Sussex	√	√	√	√	√	**Staffordshire**	√	√	x	x	x
Birmingham	√	√	√	√	√	**Abertay**	√	√	x	x	x
Exeter	√	√	√	√	√	**Heriot-Watt**	√	√	x	x	x
Southampton	√	√	√	√	√	**Strathclyde**	√	√	√	x	x
Birkbeck	√	√	√	√	√	**Aberystwyth**	√	√	x	x	x
Leicester	√	√	√	√	√	**Queen's Belfast**	√	√	x	x	x
Sheffield	x	x	√	√	√	**Durham**	√	x	√	x	x
Brunel	√	x	√	√	√	**Bucks New**	√	x	x	x	x
Aberdeen	√	√	√	√	√	**Central Lancashire**	√	x	x	x	x
City	√	√	√	√	√	**Leeds**	√	x	x	x	x
Kent	√	√	√	√	√	**SSEES**	√	x	x	x	x
Swansea	√	√	√	√	x	**Thames Valley**	√	x	x	x	x
Loughborough	√	√	√	√	x	**Ulster**	√	x	x	x	x
Dundee	√	√	√	√	x	**Bangor**	√	x	x	x	x

* In all cases, where universities have undergone name changes, or been merged with another/others, the current names are used.

Internal and external gaps and research activity

Internal gaps

Table 2 shows the relationship between the internal gap and withdrawal from the next E&E UOA with the number of universities in each category. The evidence from these data is of quite a strong relationship between the *sign* of the internal gap and the decision to withdraw, with those with a negative internal gap more likely to withdraw from the next E&E UOA than those with a positive gap. This is borne out by the fact that out of the total of 37 withdrawers, 27 had negative internal gaps and only 9 had positive internal gaps.

Table 2: Number of withdrawers from the E&E UOA by sign of the internal gap

	1992	1996	2001	2008	Totals
Positive	2	3	1	3	**9**
Zero	1	0	0	0	**1**
Negative	7	10	6	4	**27**
Totals	**10**	**13**	**7**	**7**	**37**

For some universities the decision to withdraw may be based on results from more than one REE. Table 3 shows for the universities that withdrew the number of negative internal gaps In the first column are the 11 universities that withdrew with a positive internal gap, presumably because the MARS was not achieved or for strategic reasons such as entering a different UOA, or subject changes. In the next two columns are 12 universities that withdrew on the back of one negative internal gap and in the third column are the 11 universities that withdrew after two negative gaps. In the final column is one university that only withdrew after three consecutive negative internal gaps. Why it takes some universities longer than others to withdraw is not known but it does not appear to be related to whether a university was old or new.

Table 3: The number of negative internal gaps before withdrawal

None	One	Two	Three
Swansea	Kingston*	Loughborough	Northumbria
London Met	Keele	Dundee	
Stirling	Liverpool	Manchester Met.	
De Montfort	Newcastle	Bath	
East London	Staffordshire	Hull	
Portsmouth	Aberystwyth	Nottingham Trent	
Reading	Durham*	SOAS	
Strathclyde	Central Lancashire	Salford	
Buckinghamshire New	Leeds	Abertay	
Ulster	SSEES	Heriot-Watt	
Bangor	Thames Valley	Queen's Belfast	
	Brunel		

* These universities withdrew, returned then withdrew again

Table 4 breaks down this information into old and new universities. The conclusion from this table is that old universities were more likely to withdraw after a negative internal gap: 18 out of 23 old universities (78%) withdrew while for new universities it was 9 out of 14 (64%). The corollary is that new universities were more likely to withdraw following a positive internal gap: 4 out of 14 (29%) new universities withdrew after a positive internal gap against 5 out of 23 (22%) for old universities.

Table 4: The number of withdrawers from the E&E UOA by sign of internal gap for old and new universities

	1992		1996		2001		2008		Totals	
	Old	**New**	**Old**	**New**	**Old**	**New**	**Old**	**New**	**Old**	**New**
Positive	2	0	1	2	0	1	2	1	**5**	**4**
Zero	0	1	0	0	0	0	0	0	**0**	**1**
Negative	4	3	7	3	5	1	2	2	**18**	**9**
Totals	**6**	**4**	**8**	**5**	**5**	**2**	**4**	**3**	**23**	**14**

As well as the *sign* of the internal gap there is also the *size* of the internal gap. It could be argued that the larger the negative internal gap the more likely a university is to withdraw as it means that the economics score is well below the overall university average. One of the problems in comparing the size of gaps through time is the different measures used in different REEs. This makes it more difficult to draw accurate direct comparisons of gap sizes between withdrawers and those continuing in the E&E UOA. Notwithstanding this issue, the computed average size of gap for withdrawers was -0.564 and for continuers -0.554 indicating very little difference between the two groups. This casts doubt on whether the size of gap had had any influence on the decision to withdraw, in contrast to evidence of the importance of the *sign* of the gap. Indeed some of the largest gaps (e.g. Sheffield -1.85, Surrey -1.53, Aberdeen -1.14, Bristol -1.09 and Manchester -1.07, all in 2001) did not lead to withdrawal. None of the withdrawers from the E&E UOA has given up on research completely. Each still enters REEs but in different UOAs, many having moved economists to the B&M UOA, contributing to the healthy number of universities in the B&M submission reported earlier.

External Gaps

When the external gaps are included in the analysis a much clearer picture emerges. Figure 4 is a plot for each university in the E&E UOA of the 2014 REF showing both the internal and external gaps. In quadrant 1 are nine universities that had positive internal and external gaps. These might be considered the elite economics research universities. In quadrant 3 are 15 universities that had both negative internal and external gaps. Quadrant 4 has four universities that had negative internal gaps and positive external gaps. Quadrant 3 is empty.

The significance of where universities are located in the quadrants will become clear after the following discussion of the withdrawers from previous REEs. In the figures that follow the withdrawers are the named universities that are most likely to be in quadrant 3 with a double negative gap. Only the results for the three most recent REEs are shown in the text, the two earliest are in the Appendix. Figure 5 for 2008 shows that three of the withdrawers - London Met., Stirling and Swansea had a positive internal gap and a negative external gap while Dundee, Kingston, Loughborough and Manchester Met. had double negative gaps. In Figure

6 for 2001, three universities - East London, Newcastle and Liverpool withdrew after a negative gap. A good example of double negative gaps is in Figure A2 which shows the results of the 1996 RAE. Of the thirteen withdrawers, eleven are in quadrant 3. Of the two others, one (Reading) is in quadrant 1 and the other (De Montfort) is in quadrant 2. These two universities are very much the exception to the rule. In Figure A1 for the 1992 RAE, all ten withdrawers are in quadrant 3. To summarise, over four REEs, only two of the withdrawers came from quadrant 4, five came from quadrant 2 and 30 came from quadrant 3. None of the withdrawers came from quadrant 1. The double negative gap is therefore then a good predictor of whether or not a university will withdraw from the next REE though a number of universities with double negative gaps still went on to submit to the next E&E UOA.

Figure 4: E&E internal and external gaps 2014

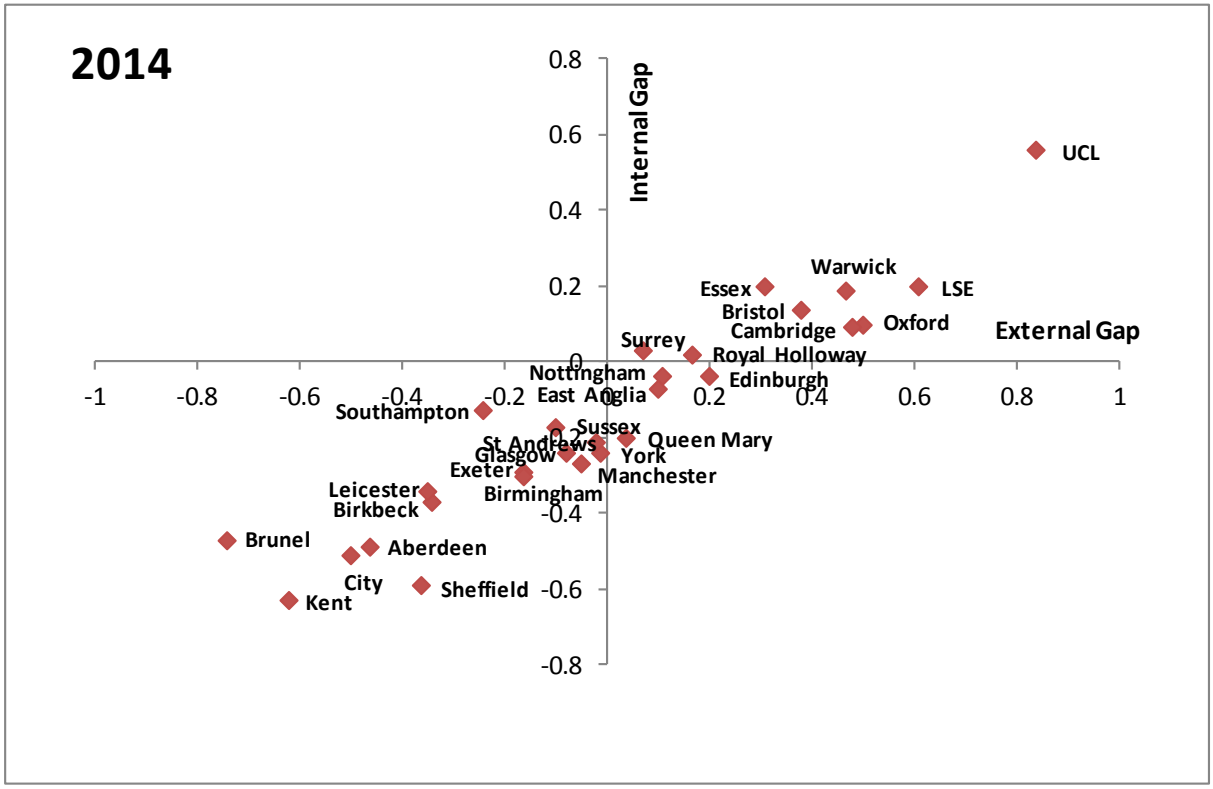


Figure 5: E&E Internal and External Gaps 2008

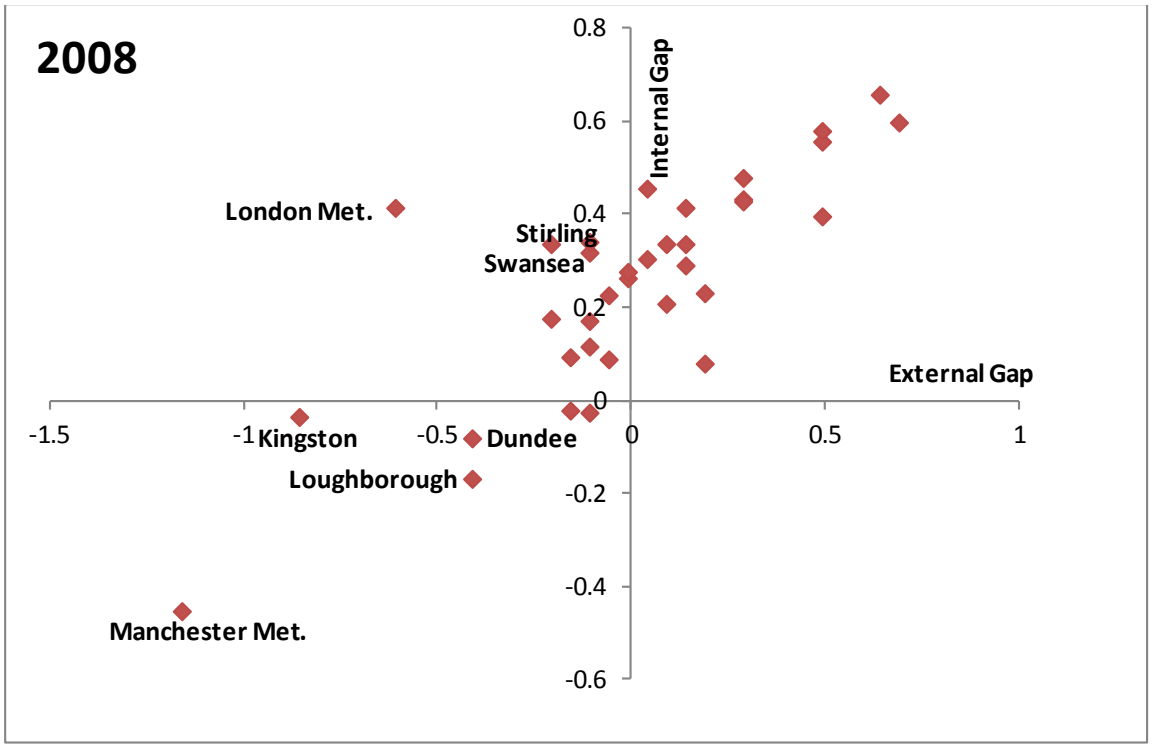
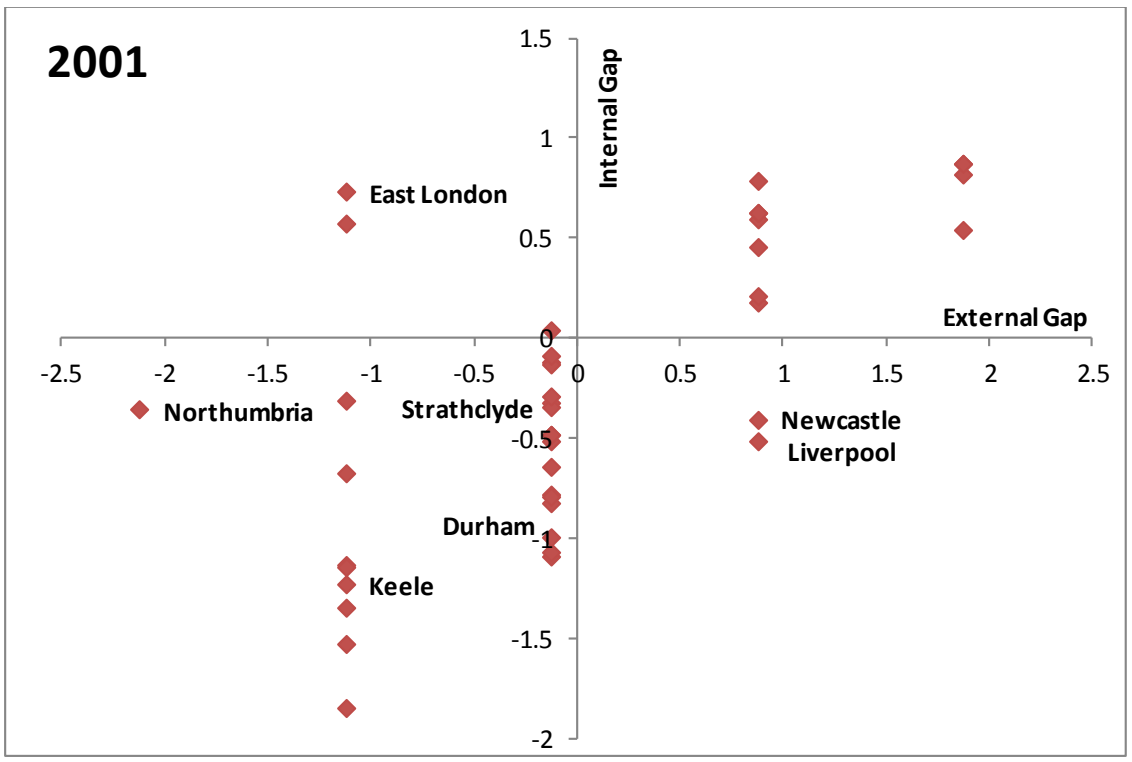


Figure 6: E&E internal and external gaps 2001



A probit analysis of the enter/leave decision

A probit regression is used when there is a binary dependent variable, that is, where the dependent variable can only take two values. Standard OLS regressions do not work well when the dependent variable is binary. The dependent variable in this model is *submit* (to the next REE) with two alternatives, stay and leave. The independent variables are the internal gap, the external gap and the product of the two. Table 5 shows the mean values of the internal and external gaps for stayers and leavers based values of 1 for a positive gap and 0 otherwise. The larger mean values for the stayers, is as expected. The value for the mean external gap for leavers is very low (0.027) because only one of the 37 leavers had a positive external gap. This meant that for the leavers, the mean values of the product is also 0.027 suggesting almost perfect collinearity between the external gap and the product, making estimation very difficult.

Table 5: The mean values of internal and external gaps at the previous REE

Gap	Stayers[1]	Leavers[2]	All
Internal	0.677	0.243	0.591
External	0.563	0.027	0.456
Product of internal and external gaps	0.516	0.027	0.419
Number of cases	149	37	186

[1] Institutions that submitted to the E&E UOA at the next REE. [2] Institutions that withdrew from the E&E UOA at the next REE.

Two approaches to modelling the leave/stay decision are adopted. The first column of Table 6 shows the results of the estimation of an exploratory linear probability model. The dependent variable (submit) is a binary variable that takes a value of 1 if an institution stays in the UOA at the next REE, 0 if it leaves. The independent variables are also binary taking the values of 1 for a positive gap and 0 otherwise. Table 6 shows that the signs of the estimates are in line with expectations in that both a positive internal and a positive external gap exert a positive and statistically significant impact on the probability of staying in the E&E UOA at the next REE. The product of the internal and external gaps is negatively signed and insignificant. This is a slightly surprising result as a double positive gap is expected to provide a strong justification for continuing with a submission. Though a useful starting point, given the well-known problems of bias and inconsistency in OLS estimates of parameter coefficients in linear probability models, a probit model was estimated.

Table 6: Submit regressed on the internal and external gaps

	Linear Probability Model	Probit Model		
Constant	0.594*** (13.07)[3]	0.366** (2.29)[4]	0.249 (1.65)*	0.238 (1.56)
Internal Gap [1]	0.156** (2.02)	1.056*** (4.66)	0.398 (1.45)	0.436 (1.53)
External Gap [2]	0.406*** (2.84)	-	1.657*** (3.84)	14.001 (0.00)
Product of the internal and external gaps	-0.170 (1.04)	-	-	-12.443 (0.00)
Number of cases	186	186	186	186

*****, **, *** significant at the 1%, 5% and 10% levels.

[1] The internal gap takes the value of 1 for a positive gap and 0 otherwise.

[2] The external gap takes the value of 1 for a positive gap and 0 otherwise.

[3] t statistics

[4] z statistics

In the first specification only the internal gap is entered as an explanatory variable. It emerges as highly significant and indicates that a positive internal gap is associated with a higher probability of remaining in the UOA at the next REE. In the second specification both the internal and external gaps are entered together. Interestingly, the result of this is to lower the size of the coefficient on the internal gap and though it retains the same sign it loses its statistical significance. In contrast, the external gap not only has the expected sign but is strongly significant. These results suggest that what carries greater weight in the decisions on whether to continue to submit to a UOA or not is not how a unit has done relative to its internal peers but how it has performed relative to other institutions. It appears to be the case that external competition is more important than internal competition in internal resource allocation decisions. The third specification includes the product and external gap variables. The high level of collinearity between these variables just referred to render the results of little value. It is included for completeness.

Discussion

Referring back to Figure 4 for the 2014 REF, is it possible to predict the likely withdrawers from the next REE, whenever it comes round? Universities that did poorly in the 2014 REF

might be vulnerable just as in previous years some poor performers have dropped out. But what does poor performance actually mean? Does it mean the same for all universities or does it mean different things to different universities. To a university with a strong research pedigree, a top five or ten elite research institution, anything less than a score that gives a position in the top elite would be disappointing, to say the least. Elite universities are likely to have a MARS that reflects their high status and position. Individual research groups will also look at other research groups in their own university with a view to outdoing them in the REE. As scores increasingly determine future funding, under a framework of research excellence, it seems likely that funding will become more and more concentrated in fewer universities. Already funding is highly concentrated towards elite institutions (Lee, Pham and Gu, 2013) and is likely in the future to exclude marginal performers altogether. In the light of this move the reluctance of universities to commit resources to research of questionable (RAE/REF–able) quality is understandable. It then makes it more likely that, just as happened in the past, some universities will pull out of the E&E UOA at the next REE.

Such universities may be those that do not meet their MARS, where researchers in other UOAs in their own universities do better or those that perform worse than fellow economists in other universities. Certainly not at risk of pulling out for reasons of poor performance are those elite universities in quadrant 1 of the plots. In the UK higher education landscape these institutions have held a lofty position ever since the start of REEs. Here, upward and downward mobility is virtually absent. The top five or six have pretty much always been the top five or six. They have always had a double positive gap, performing better than the average of researchers in their own universities and better than the average of economics researchers in the E&E UOA. At the bottom of the table are those that performed poorly in the 2014 REF.

Some of these fifteen institutions with a double negative gap may be particularly vulnerable to withdrawing from the E&E UOA. As Table 7 shows, seven are in the *Russell Group*, and eight are not currently in a named group especially since the demise in November 2013 of the *1994 Group.* In line with the view that universities may be more likely to withdraw following more than one double negative gap the final column shows, for those that had a double negative gap in 2014, the earlier occurrences of double negative gaps. At risk then according to this approach would be four of the 'Others' group - St Andrews, Kent, City and Brunel. Elsewhere Manchester and Leicester have both had three double negative gaps.

Table 7: Universities' positions on the REF2014 research graph

University	Q1 (positive internal and external gaps) = 'double positive'	Q2 (positive internal gap and negative external gap)	Q3 (negative internal and external gaps) = 'double negative'	Q4 (negative internal gap and positive external gap)	Others REEs of double negative for those that were double negative in 2014
Russell Group (16)					
Birmingham			√		2001
Bristol	√				
Cambridge	√				
Edinburgh				√	
Exeter			√		
Glasgow			√		2001
LSE	√				
Manchester			√		1992, 2001
Nottingham				√	
Oxford	√				
Queen Mary				√	
Sheffield			√		2001
Southampton			√		
UCL	√				
Warwick	√				
York			√		2008
Others (12)					
Birkbeck			√		
East Anglia				√	
Essex	√				
Leicester			√		1992, 1996
Royal Holloway	√				
Sussex			√		
Aberdeen			√		2001
Brunel			√		1992, 2001
City			√		1992, 2001
Kent			√		1992, 2001
Surrey	√				
St Andrews			√		1992, 2001, 2008
Totals (28)	***9***	***0***	***15***	***4***	

Much of this analysis of the gaps depends on how the gaps are calculated. This in turn depends on whatever is used as the minuend and subtrahend. The minuend is the given research score in the E&E UOA so there should be little debate about it. In both the internal and external gaps the subtrahend is an FTE weighted mean value. The construction of the internal gap has a strong intuitive appeal because it compares for each university the E&E score with average of all UOAs in that university. With the external minuend, the average of E&E scores for all universities might be the wrong one to use. The elite universities will be well aware who they wish to better or emulate. They will not be overly concerned with how those at the lower end of the scale fare. The proxy for the MARS used here is one of several that could be used. For continuers in a UOA the most recent score is known to be enough to keep a university in a UOA so it is at least as high as the MARS, but, it may be amended as

a consequence of shifts in strategy, changes in research staff through attrition and retirements, and funding restrictions. Using different minima for different groups of universities is another option. The sizes of the negative gaps do not appear to have a strong influence on the decision to withdraw but if they do, those furthest to the south-west - Kent, City, Sheffield, Brunel and Aberdeen - might be especially vulnerable irrespective of the number of double negative gaps they have recorded in the past.

Conclusions

In the UK and many other countries research assessments are periodically carried out by the state to measure the quality of research to ensure that state funding goes to the best researchers. So far in the UK there have been five fully-fledged research evaluation exercises going back to the 1990s. In this paper we focus on the evaluations carried out in 1992, 1996, 2001, 2008 and 2014 and ignore the first two preliminary exercises in 1986 and 1989 which globally, were probably the first formal attempts to assess research quality (Bence and Oppenheim 2005). From the beginning, universities have entered their research outputs into different subject areas or (UOAs) for assessment by their peers. They have not relied on where the research is published as a measure of its quality. Since 1992 universities have decided what areas to support in each subsequent REE based on results at the previous REE. As part of the process of learning what they were good at and not so good at, universities entered new UOAs in the hope of achieving high scores and withdrew from other UOAs largely because of poor performance consistent with the models presented earlier. One UOA from which universities were especially likely to withdraw was E&E. In 1992, 60 universities entered this UOA and by 2014 the number had fallen to just 28 with no post-1992 universities entering in 2014. This was one of the highest percentage withdrawal rates of any UOA. STEM subjects such as the Physics, Chemistry, Biological Sciences and Civil Engineering UOAs followed similar paths with substantial declines in submissions. The largest rises were for English Language and Literature; and Business and Management.

In attempting to explain the withdrawal of so many universities from the E&E UOA a new concept of the MARS was discussed in relation to the decisions made by universities to allocate resources to research activity. The focus was on the REEs and the decision to withdraw from the E&E UOA. This UOA is discussed in detail and the decline in the number of universities entering this UOA is documented over the course of five REEs. A withdrawal model is devised and tested. The model is based on how well a university performed on two counts, first, in the E&E UOA relative to other UOAs in the same university and second, in the E&E UOA relative to other universities in the E&E UOA. Two forms of gap were devised. The internal gap compares the performance in the E&E UOA with the average for the same university. A positive internal gap indicates that E&E has performed better than an FTE weighted average of all UOAs in that university and a negative internal gap that it has performed worse than the average. The second gap is the external gap. This shows how the performance of E&E in a university compares with an FTE weighted average score of all universities in the E&E UOA. A positive external gap shows that a university has performed better than the average of all universities in the E&E UOA while a negative gap shows it has performed worse. The data were plotted in graphs showing all four quadrants, the most telling being quadrant 3 which contained universities with a double negative gap. There was strong evidence that double negative gaps were a good predictor of the withdrawal of universities from the E&E UOA. A probit analysis showed that the external gap was more

important than the internal gap in decisions on whether to continue to submit to the E&E UOA. It is thus not how well a unit has done relative to its internal peers but rather how it has done against other institutions that really matters.

It will be interesting to see if the model stands up when predicting withdrawal from other UOAs. The report that the University of Surrey intended to close its Politics department with substantial job losses after a very poor performance in REF2014, 100% student satisfaction with Politics programmes, is one example of the power of research in UK universities (Field, 2015). If the model presented in this paper is robust it will not be UOA-specific and should apply just as well to other subjects.

References

Agyemang, G. and Broadbent J. (2015) Management control systems and research in universities. *Accounting, Auditing & Accountability Journal*, 28 (7), 1018-1046.

Bence, V. and Oppenheim, C. (2005) The evolution of the UK's Research Assessment Exercise: publications, performance and perceptions. *Journal of Educational Administration and History,* 37 (2), 137-155.

Curran, P. (2000) Competition in UK higher education. *Higher Education Quarterly,* 544 (4), 386-410.

Field, J. (2015) Surrey uses REF results to justify job cuts. *Research Fortnight*, 25 March.

Grove, J. (2015) Newcastle University staff express fears over new 'targets'. *Times Higher Education*, December 17.

Guena, G. and Martin B. R. (2003) University research evaluation and funding: an international comparison. *Minerva*, 41, 277-304.

Harley, S. and Lee, F. S. (1997) Research selectivity, managerialism, and the academic labour process: the future of non-mainstream economics in U.K. universities, *Human Relations*, 50 (22), 1427-1460.

Hicks, D. (2012) Performance-based university research funding systems. *Research Policy*, 41, 251-261.

Jarratt, Sir Alex (chair) (1985) *Report of the Steering Committee for Efficiency Studies in Universities* (The Jarratt Report), report to the Committee of Vice-Chancellors and Principals (CVCP) and to the University Grants Committee (London: CVCP).

Johnston, J., Reeves, A. and Talbot, S. (2014) Has economics become an elite subject for elite UK universities? *Oxford Review of Education*, 40 (5), 590-609.

Lee, F. S. (2007) The research assessment exercise, the state and the dominance of mainstream economics in British universities. *Cambridge Journal of Economics,* 31, 309-325.

Lee, F. S., Pham, X. and Gu, G. (2013) The UK research assessment exercise and the narrowing of UK economics. *Cambridge Journal of Economics,* 37, 619-773.

Leisyte, L. and Westerheijden, D. (2014) Research evaluation and its implications for academic research in the United Kingdom and the Netherlands. Discussion papers des Zentrums für HockschulBildung Technische Universität Dortmund 01-2014.

Lucas, L. (2006) *The Research Game in Academic Life*. Maidenhead: Open University Press.

Martin, B. and Whitley, R. (2010) The UK Research Assessment Exercise: A case of regulatory capture? In: R. Whitley, J. Glaser and L. Engwall (eds.) *Reconfiguring Knowledge Production: Changing Authority Relationships in the Sciences and their Consequences for Intellectual Innovation.* Oxford: Oxford University Press, 51-80.

Merton, R. K. (1968) The Matthew effect in science. *Science,* 159 (3810), 56-63.

Mingers, J. and Willmott, H. (2013) Taylorizing business school research: on the 'one best way' performative effects of journal ranking lists. *Human Relations*, 66 (8), 1051-1073.

Owens, B. (2013) Research assessments: judgement day. *Nature,* 502 (1741).

RAE 1992 (1992) Research Assessment Exercise 1992: The Outcome. Available at: http://www.rae.ac.uk/1992/index.html (Accessed 3.9.15).

RAE 1996 (1996) 1/96 1996 Research Assessment Exercise: The Outcome (December 96). Available at: http://www.rae.ac.uk/1996/index.html

RAE 2001 (2001) Results. Available at: http://www.rae.ac.uk/2001/results/ (Accessed 2.10.14).

RAE 2008 (2008) Quality Profiles. Available at: http://www.rae.ac.uk/results/ (Accessed 12.10.14).

Rafols, I., Leydesdorff, L., O'Hare, A., Nightingale, P. and Stirling, A. (2012) How journal rankings can suppress interdisciplinary research. *Research Policy*, 41, 1262.

REF2014 (2014) Results and Submissions. Available at: http://results.ref.ac.uk/ (Accessed 18.12.14).

Russell Group (2009) The concentration of research funding in the UK: driving excellence and competing globally. Available at: http://www.russellgroup.ac.uk/uploads/Concentration-of-research-funding.pdf (Accessed 18.3.15).

Sato, I. and Endo, T. (2014) From the RAE-able to the REF-able? A note on formative reactivity in national research quality assessment. *Research on Academic Degrees and University Education*, 16 (November), 85-104.

Sharp, S. (2004) The research assessment exercises 1992-2001: patterns across time and subjects. *Studies in Higher Education*, 29 (2), 201-218.

Thomas, E. (2007) National research assessment in higher education. In H. De Burgh, J. Black and A. Fazackerly (eds.) *Can the Prizes Still Glitter? The Future of British Universities in a Changing World.* Buckingham: University of Buckingham Press, 39-47.

Willmott, H. C. (2011) Journal list fetishism and the perversion of scholarship: reactivity and the ABS list. *Organization,* 18 (4), 29-44.

Appendix

Figure A1: E&E internal and external gaps 1992

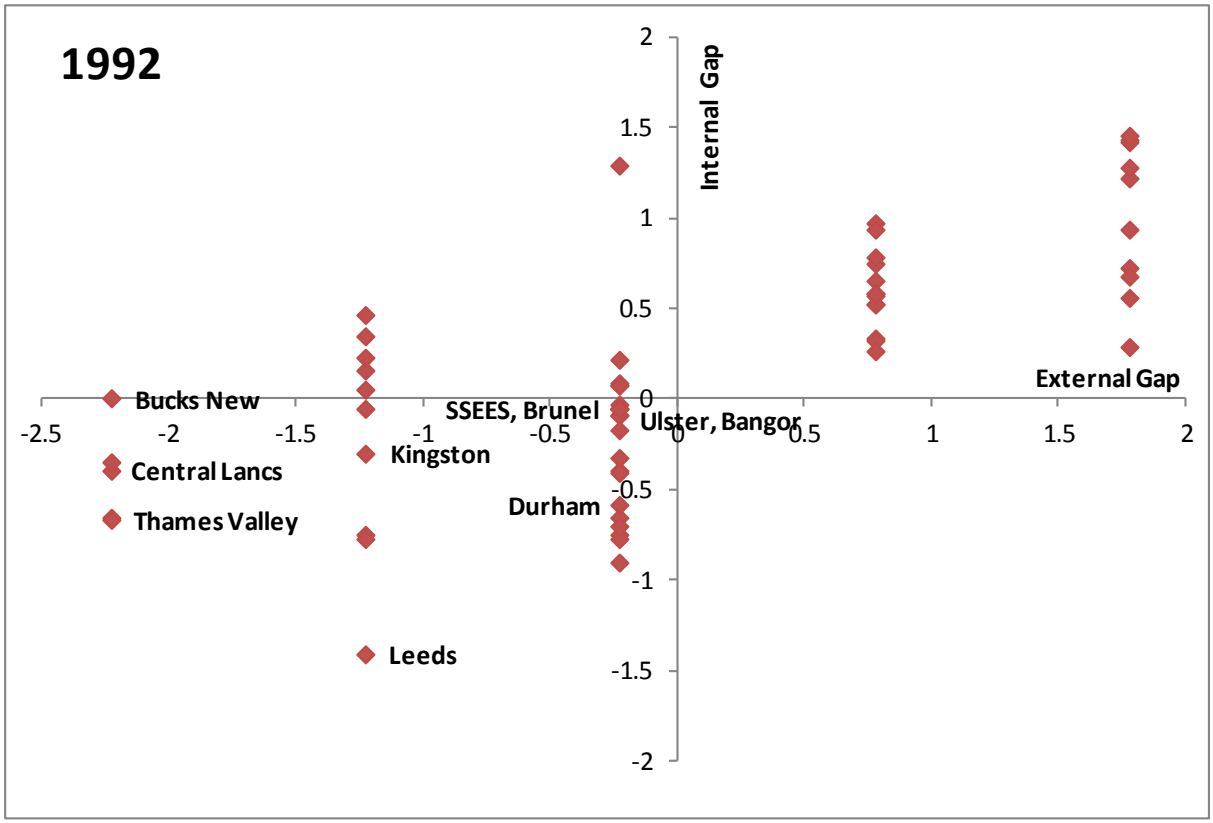


Figure A2: E&E internal and external gaps 1996